Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
William W. Castillo Guardado (State Bar No. 294159)
Holden J. Benon (State Bar No. 325847)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                czirpoli@saverilawfirm.com
                cyoung@saverilawfirm.com
                lkessler@saverilawfirm.com
                eabuchanan@saverilawfirm.com
                wcastillo@saverilawfirm.com
                hbenon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@butericklaw.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*

[Additional counsel on signature page]

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

J. DOE 1 et al., *Individual and Representative Plaintiffs,* v. GITHUB, INC., et al., *Defendants.*	Case Nos. 4:22-cv-06823-JST 4:22-cv-07074-JST **JOINT CASE MANAGEMENT STATEMENT** Date: March 22, 2024 Time: 1:30 p.m. Courtroom: Videoconference Judge: Hon. Jon Tigar

Case No. 4:22-cv-06823-JST

JOINT CASE MANAGEMENT STATEMENT

1 Plaintiffs Does 1 through 5 and the proposed class ("Plaintiffs"); Defendants GitHub, Inc.

2 ("GitHub"); Microsoft Corporation ("Microsoft"); and OpenAI, Inc.; OpenAI, LP; OpenAI OpCo, LLC;

3 OpenAI GP, LLC; OpenAI Startup Fund GP I, LLC; OpenAI Startup Fund I, LP; OpenAI Starup Fund

4 Management LLC, OpenAI LLC, OAI Corporation, OpenAI Holdings LLC, OpenAI Holdco LLC,

5 OpenAI Investment LLC, Open AI Startup Fund SPV I, LP, and OpenAI Startup Fund Management, LLC

6 ("OpenAI") file this Joint Case Management Statement ("CMS") pursuant to ECF No. 186.

7 **I.   STATUS OF DISCOVERY**

8   **A.   Plaintiffs' Discovery Requests**

9   **1.   Discovery Propounded by Plaintiffs and Responses from Defendants.**

10 **Requests for Production of Documents ("RFPs"):**

11 **Plaintiffs' Position:** On July 12, 2023, Plaintiffs served First Sets of Requests for Production on

12 GitHub (49 requests), Microsoft (60 requests), and OpenAI (57 requests). Defendants served responses

13 and objections on August 18, 2023. The First Set of Requests was served prior to the Court's entry of the

14 ESI protocol, which provides for the discovery of electronically stored information. In the September 8,

15 2023, Joint Case Management Statement (ECF No. 153), Defendants attributed their failure to produce

16 any documents to the lack of an ESI Protocol in this case. Then, Defendants ignored repeated requests to

17 meet and confer on Plaintiffs' draft ESI Protocol until shortly before the deadline set by this Court. See

18 ECF No. 166. The Court entered a version of the Northern District of California ESI Patent Protocol

19 which, among other things, limits ESI discovery substantially.

20   On March 4, 2024, Plaintiffs served their second set of requests for production to Defendants

21 Microsoft and GitHub. On March 14, 2024, Plaintiffs served their third request for production of

22 documents, emails specifically, pursuant to the ESI protocol, on Defendants Microsoft and GitHub. The

23 parties have yet to respond.

24   As a general matter, Plaintiffs have sought information regarding corporate structure; identification

25 of those with management or other responsibility for the conduct alleged; the background and operation of

26 the relevant technologies; the source and use of the training data and datasets; use of and compliance with

27 applicable licenses; matches or near matches of plaintiffs' data; submissions of information to regulators

28 and governmental agencies; public statements regarding these topics; business and financial relationships

between and among the parties and their principals; income, revenue and profits from sale or distribution of Copilot; and insurance policies and other materials ordinarily produced in cases of this type.[1] Broadly speaking, Defendants objected to all of Plaintiffs' first set of RFPs. Microsoft and GitHub first produced documents on December 12, 2023. The productions consisted of 22 pages of Microsoft documents and approximately 300 GitHub documents. OpenAI has produced 39 documents—almost all publicly available.

Disputes remain as to the first set of document requests and negotiations with Defendants continue. OpenAI continues to stall in producing the training data for Codex/Copilot and the Parties will continue to negotiate the terms of access to this training data. Discovery regarding the training data, including the identification of the sources of the training data and its use by Defendants in the products at issue in this case is a threshold discovery issue.

Plaintiffs have conducted face-to-face meet and confer sessions with GitHub, Microsoft, and OpenAI on the remaining first set of discovery requests Plaintiffs have propounded thus far, conferring for hours with each Defendant.[2] The parties have continued to discuss specific requests by letter and Plaintiffs have requested dates for further meet and confer sessions from Defendants. Plaintiffs and OpenAI met and conferred by video call on December 13, 2023. Consistent with the principles of cooperation underlying discovery as outlined by the Federal Rules, the Sedona Conference principles, and the Northern District of California's Guidelines for the Discovery of Electronically Stored Information, the Parties have met and conferred with respect to Defendants' responses numerous times.

Generally, Plaintiffs anticipate Defendants will continue to produce responsive documents through (1) collection of responsive materials within their possession, custody, or control (*i.e.*, "go and get") and (2) implementing search terms for custodial ESI. Contrary to Defendants' assertions, Plaintiffs' proposed custodians and search terms served on Microsoft and GitHub (February 29, 2024) and OpenAI (March 14, 2024) adhere to the ESI protocol because they relate to *non-email discovery* relevant to Plaintiffs' First Set

---

[1] This description is intended to generally describe the nature and scope of the document discovery sought by Plaintiffs to date. Plaintiffs and Defendants are meeting and conferring on the individual requests. Plaintiffs hope to bring this process to a close as soon as possible.

[2] Plaintiffs met and conferred with respect to Microsoft and GitHub separately and individually for a total of approximately five to six hours.

1   of Requests for Production on all Defendants. The ESI Protocol does not limit the number of custodians

2   and search terms Plaintiffs can request for non-email discovery. Pursuant to the ESI Protocol, certain

3   responsive documents will be produced in native form. To date, after this Court entered Defendants' ESI

4   Protocol, Microsoft and GitHub still have not produced any documents, even those promised in their

5   respective responses to Plaintiffs' requests.

6   Plaintiffs propounded interrogatories months ago to its selection of custodians and search terms

7   which Defendants' have largely improperly refused to answer. *See In re Folding Carton Antitrust Litig.*,

8   76 F.R.D. 417, 419 (N.D. Ill. 1977) (describing the types of information requested by Plaintiffs here as

9   "classic first-wave discovery"). OpenAI has produced only a tiny fraction of the documents requested and

10   only a smattering of the documents promised. Considering Defendants' continued refusal to comply with

11   their discovery obligations, Plaintiffs anticipate a number of remaining discovery issues will require court

12   resolution, which they will bring to Magistrate Judge Ryu in due course pursuant to her recent order

13   delaying resolution of discovery disputes until after the pleadings have been resolved.

14   **OpenAI's Position:** To date, Plaintiffs have not raised any substantive complaints with OpenAI

15   regarding its August 18, 2023, responses to Plaintiffs' First Set of RFPs. Plaintiffs first informed OpenAI

16   that they wanted to discuss OpenAI's responses during a telephonic conference on December 13, 2023,

17   almost four months later. OpenAI acquiesced to Plaintiffs' demand to meet and confer before the

18   holidays, despite limited availability, and agreed to a 30-minute telephonic conference on December 20,

19   2023. OpenAI's counsel promptly joined the call, as did only one of Plaintiffs' counsel. However,

20   Plaintiffs' counsel was unprepared and unwilling to discuss any of their issues regarding OpenAI's RFP

21   responses without other Plaintiffs' attorneys present. OpenAI's counsel patiently waited for 24 minutes

22   before another Plaintiff's counsel joined the call, who immediately asked to reschedule the meet and

23   confer over email given how little time remained. Plaintiffs have yet to reach out to OpenAI to reschedule

24   this meet and confer.

25   In response to Plaintiffs' RFPs, OpenAI has produced non-custodial documents regarding its

26   corporate structure and corporate formation; employee organization charts reflecting those with

27   management or other responsibility for the conduct alleged; documents describing the operation of the

28   relevant technologies; and insurance policies applicable to this case. OpenAI also has been diligently

1   working to make the data used to train models used for Codex available for inspection. To do so, OpenAI

2   must investigate to identify the specific data used to train certain models, find and test various

3   technological solutions for making voluminous and highly confidential data available for review and

4   inspection, and prepare a protocol describing how the data can be inspected on a secured computer. Given

5   the voluminous amount and format of the data at issue, this has been a time- and labor-intensive process.

6   Plaintiffs' counsel is aware of this complicated process from other co-pending litigations they have filed

7   against OpenAI in which OpenAI has prepared and circulated a draft training data inspection protocol for

8   their review.

9   For custodial documents, OpenAI finally received Plaintiffs' first set of proposed search terms and

10   custodians on the evening of March 14, 2024—one day before the filing of this case management

11   statement. Plaintiffs' proposal blatantly disregarded the mandatory procedures set out in the Court's

12   November 1, 2023 Order Re: Discovery of Electronically Stored Information ("ESI Order"). Instead of

13   identifying five custodians and providing five search terms per custodian, Plaintiffs have proposed ten

14   custodians and twenty-one, indiscriminate and sweeping search terms, including "Codex" and "Copilot,"

15   for what appears to be all ten custodians in direct contravention of the entered ESI Order. Plaintiffs have

16   also failed to serve specific email production requests, consistent with the ESI Order.

17   **Microsoft and GitHub's Position:** The Parties have met and conferred repeatedly regarding

18   Plaintiffs' first set of requests for the production of documents and have begun producing documents.   On

19   February 29, 2024, GitHub produced a sample of its telemetry data to aid in the ongoing conferral

20   regarding feasible sampling and search methodologies of the vast amount of GitHub telemetry data.

21   On March 04, 2024, Plaintiffs served a second set of document requests on Microsoft and GitHub.

22   Microsoft and GitHub will serve objections and responses on April 03, 2024 as required by the Federal

23   Rules of Civil Procedure.

24   On November 16, 2023, Microsoft and GitHub sent a letter to Plaintiffs requesting, *inter alia*, that

25   Plaintiffs serve custodial document requests and identify specific custodians, search terms, and time frame

26   constraints as required by paragraphs 6 and 8 of the ESI Order issued in this matter. *See* ECF No. 168

27   (ESI Order). On February 29, 2024, Plaintiffs served a letter on Microsoft and GitHub proposing

28   custodians and search terms for custodial discovery. On March 13, 2024, Microsoft and GitHub

1   responded that Plaintiffs' letter does not adhere to the procedure for custodial discovery set forth in the

2   Court's ESI Order. On March 14, 2024, Plaintiffs propounded email document requests, which GitHub

3   and Microsoft are reviewing. Microsoft and GitHub will provide objections and responses when due, and

4   will engage in meet and confer with Plaintiffs regarding these new requests.

5   There are no disputes ripe for Court intervention.

6   **2.   Third Party Discovery**

7   **Plaintiffs' Position:** Beginning in September 2023, Plaintiffs propounded discovery to certain

8   investment firms who were involved in providing investments to GitHub and were involved in the

9   purchase of GitHub by Microsoft. Those requests seek information regarding the open-source licenses at

10 issue in this lawsuit as well other information regarding the transactions.

11   Plaintiffs have served the following entities with subpoenas to produce documents: (1) Morgan

12 Stanley; (2) Hydrazine Capital; (3) Andreesen Horowitz, AH Capital Management LLC; and (4) Sequoia

13 Capital (collectively, "Subpoena Recipients"). The Subpoena Recipients each provided responses and

14 objections to the Subpoenas, which Plaintiffs served on Defendants. Plaintiffs have met and conferred

15 with Morgan Stanley, Andreesen Horowitz, and Sequoia Capital.

16   As confirmed by third parties, such information is contained in investment decks and other

17 material prepared in connection with analyzing the proposed transactions. Such materials, prepared by

18 experienced sophisticated outside investors contain admissions of Microsoft and will likely shed light on

19 many aspects of the claims at issue in this case. Such information will be relevant among other things, as

20 proof of the elements of the breach of contract claim, damages and other remedies as well proof of the

21 DMCA claim. This evidence will show Defendants' knowledge and intent with regard to the fact and

22 terms of the open-source licenses and GitHub's Policies not to sell Plaintiffs' code. With respect to

23 remedies, this discovery will also serve to prove the code's commercial value for purposes of providing

24 restitution for breach of contract.

25   Neither the Defendants nor the third parties moved to quash the subpoenas. Plaintiffs met and

26 conferred over a period of several months. After Plaintiffs had negotiated and agreed to the scope of

27 production with Morgan Stanley and Andreesen Horowitz, Microsoft intervened with the third parties at

28 the eleventh hour, after agreements for production had been reached, and obtained the third parties'

1 agreement not to provide documents. With respect to one investor, Morgan Stanley, Plaintiffs' meet and

2 confer efforts resulted in an agreement to produce a limited subset of the material. On the eve of

3 production, Microsoft and GitHub intervened, preventing Morgan Stanley from producing the agreed

4 upon material. Months after the subpoena was issued, Microsoft and GitHub filed a unilateral motion for

5 protective order with Judge Ryu.[3] The parties were then ordered to meet and confer. Having done so, a

6 joint letter brief was filed on February 23, 2024. This meeting resulted in a joint discovery letter and an

7 order delaying resolution until after Defendants' motions to dismiss the Second Amended Complaint

8 (ECF No. 200) are resolved. *See* ECF No. 226.

9   Microsoft takes the position that Plaintiffs must withstand all legal challenges to their claims in

10 order to obtain such discovery. Further, Microsoft makes the additional argument that Plaintiffs are only

11 entitled to information regarding damages once the court has certified the class. The Court has sustained

12 the breach of contract claims against Microsoft and GitHub. Microsoft and GitHub dispute the existence

13 of enforceable contracts and dispute Plaintiffs' interpretations of the licenses, terms of service and other

14 written agreements. Microsoft and GitHub refuse to produce material until the Court adjudicates

15 Plaintiffs' theories of breach of contract, including the existence of the contracts, elements of a contract

16 claim, availability of alternatives of liability, damages and remedies. Plaintiffs contend that the discovery

17 rules specifically and the Federal Rules of Civil Procedure more generally do not so provide. Plaintiffs

18 also contend the Rule 26 proportionality provisions broadly support this discovery given the resources of

19 the parties, the amount of the dispute, the importance of the litigation and the relevance of this discovery.

20 Plaintiffs also contend that the merits and class discovery has not been bifurcated, and therefore, Plaintiffs

21 need not await a class certification determination in order to obtain discovery.

22   **Microsoft and GitHub's Position:** Several letter briefs have been submitted to Judge Ryu

23 regarding the propriety of these subpoenas, including by third parties and by Microsoft, GitHub, and

24 Plaintiffs. As explained more fulsomely in those briefs, Microsoft's Acquisition of GitHub (the

25 "Acquisition") is irrelevant to Plaintiffs' remaining breach of contract claims. The Acquisition is between

26 Microsoft and GitHub; it has nothing to do with the terms of contracts agreed between GitHub and its

27

28 [3] Plaintiffs had sought similar material directly from Microsoft and GitHub. Both objected and refused to produce it.

1   customers, or between Microsoft and OSS licensors, whether any GitHub Copilot output matches

2   Plaintiffs' code, whether any such output has in fact caused reputational harm, or how to ascertain the

3   amount of injury, if any, caused to each of the Plaintiffs. Moreover, motive or intent is also irrelevant to

4   breach of contract claims. The Acquisition simply has nothing to do with any of these elements of the

5   claims. Rather, Plaintiffs seek information regarding the Acquisition because they believe the valuation of

6   GitHub, Inc. is somehow a proxy for a classwide damages number. Judge Ryu denied all of the letter

7   briefs and ordered the parties to file another letter brief arguing the relevance of the Acquisition after the

8   pleadings are settled. ECF. 226.

9   Further, Plaintiffs' allegation that Microsoft intervened and obtained agreements from third parties

10   not to produce documents is patently false. The third parties moved to prevent production on their own

11   accord. As discussed below in the "Issues Requiring the Court's Attention" section, Microsoft and

12   GitHub request bifurcation of classwide damages discovery to put a stop to Plaintiffs' relentless campaign

13   into discovery that can only be relevant to classwide damages, including these third-party subpoenas.

14   **Interrogatories:**

15   **Plaintiffs' Position:** On July 19, 2023, Plaintiffs propounded 14 interrogatories each on GitHub,

16   Microsoft, and OpenAI. The interrogatories, among other things, ask Defendants to identify persons

17   responsible for the subject matter or events relevant to Plaintiffs' claims, description or identification of

18   training data, and identification and location of current and former employees who are likely to possess

19   relevant information. These interrogatories are routine first-wave discovery requests propounded in nearly

20   every case. None of Plaintiffs interrogatories could require disclosure of Plaintiffs' true names to anyone.

21   Following an extension, on August 25, 2023, Defendants provided limited responses to some

22   interrogatories, indicating their investigation was not complete, and objected to others. Initial meet and

23   confer sessions have been conducted for all outstanding interrogatories propounded by Plaintiffs. Since

24   then, the parties have continued to communicate about the interrogatory responses by letter. Magistrate

25   Judge Ryu ordered OpenAI to provide verifications for its interrogatory responses, which it did. Plaintiffs

26   requested the same from Microsoft and GitHub. Plaintiffs have had further face-to-face meet and confer

27   sessions with Microsoft and GitHub related to the interrogatories regarding Microsoft's acquisition of

28   GitHub. But they refused to produce such information as described above, instead objecting to providing

1   any information about the GitHub acquisition and objecting at the eleventh hour to subpoenas issued to

2   third parties involved in the acquisition. Defendants have not supplemented or amended their initial

3   responses. Considering Defendants' continued refusal to provide any more than their initial inadequate

4   responses, Plaintiff anticipate bringing disputes regarding their interrogatory responses to Magistrate

5   Judge Ryu.

6   **Microsoft and GitHub's Position:** Contrary to Plaintiffs' assertion that Defendants have not

7   supplemented their initial responses, GitHub served supplemental responses on December 13, 2023. Both

8   Microsoft and GitHub served verification of their respective responses. Besides the interrogatories

9   regarding the Acquisition, Plaintiffs have not raised further disputes with Microsoft and GitHub's

10 responses.   There are no disputes ripe for Court intervention.

11 **OpenAI's Position:** Contrary to Plaintiffs' assertion that Defendants have not supplemented or

12 amended their initial responses, OpenAI served amended interrogatory responses on January 19, 2024,

13 after meeting and conferring with Plaintiffs on December 13, 2023. These responses identified current

14 and former board members of OpenAI, Inc. and employees who have had responsibility for the relevant

15 technical aspects of Codex. Since then, Plaintiffs have not raised any further disputes with OpenAI's

16 responses.

17 **B.   Defendants' Discovery Requests**

18 **1.   Discovery propounded by Defendants and responses from Plaintiffs.**

19 **Requests for Production of Documents ("RFPs"):**

20 **Plaintiffs' Position:** On July 12, 2023, OpenAI served a separate set of RFPs on each of the five

21 plaintiffs. On August 11, 2023, all Plaintiffs served objections and responses to OpenAI's first sets of

22 RFPs and produced responsive documents (131 pages). Plaintiffs' production is complete. On July 27,

23 2023, GitHub and Microsoft jointly served separate sets of RFPs on each of the five Plaintiffs, broadly

24 seeking material related to threats that Plaintiffs received (the "Joint RFPs"). On August 28, 2023,

25 Plaintiffs served objections and responses to the Joint RFPs. Specifically, Plaintiffs objected and refused

26 to produce in response to Requests 2 through 5 on the grounds that the RFPs seek information protected

27 by attorney client privilege; are duplicative of OpenAI's RFPs and interrogatories; and have been the

28 subject of law and motion practice and resolved by the Court (*See* ECF Nos. 71, 77, 77-1, 77-2, 77-3, 77-

1   4, 77-5, 77-6, and 77-7). Plaintiffs are prepared to meet and confer on their responses should Defendants

2   ask them to.

3   To the extent Defendants accuse Plaintiffs of deleting relevant ESI, the relevant material is either

4   preserved or in Defendants' care, custody, and control. After meeting and conferring on this issue,

5   Plaintiffs provided revisions to Defendants' proposed stipulation on this issue but have not received a

6   response.

7   **OpenAI's Position:** OpenAI intends to meet and confer with Plaintiffs regarding Plaintiffs'

8   deficient production, which includes unreadable JSON files, a blank HTML file, emails without any

9   metadata, and improperly redacted PNG files of Slack and other messages. At this juncture, these disputes

10   do not require Court intervention.

11   **Microsoft and GitHub's Position:**   Microsoft and GitHub agree with OpenAI that Plaintiffs'

12   production is incomplete and deficient. It will meet and confer with Plaintiffs. At this juncture, these

13   disputes do not require Court intervention.

14   **Interrogatories:**

15   **Plaintiffs' Position:** On July 12, 2023, OpenAI served a separate set of interrogatories on each of

16   the five Plaintiffs. On August 11, 2023, Plaintiffs served objections and complete responses to OpenAI's

17   interrogatories.

18   On October 11, 2023, GitHub propounded interrogatories on the Plaintiffs. All Plaintiffs provided

19   complete responses on November 13, 2023. On March 5, 2024, GitHub and Microsoft sent Plaintiffs a

20   letter requesting that Plaintiffs respond more fully to Interrogatories 1-10, and provide largely irrelevant

21   information or information already in their possession. GitHub requests Plaintiffs identify their activity on

22   GitHub, that they identify all accounts used for personal communications, all devices used to create code

23   posted to GitHub, the threats Plaintiffs have received, repositories in GitHub Plaintiffs have deleted, the

24   "original" code on GitHub that Plaintiffs claim they have a copyright to, a description of their use of

25   Copilot, "identification of every communication in which plaintiffs disclosed their involvement in this

26   action," and identification of all harm Plaintiffs allege to have suffered.

27   The Parties continue to negotiate the scope of the responses to these Interrogatories.

28

1 **OpenAI's Position:** On December 19, 2023, OpenAI sent Plaintiffs a letter notifying them of

2 deficiencies in their interrogatory responses. After OpenAI and Plaintiffs met and conferred on January 4,

3 2024 regarding these deficiencies, Plaintiffs served supplemental responses on January 19, 2024. On

4 February 8, 2024, OpenAI identified additional deficiencies with Plaintiffs' supplemental responses to

5 Plaintiffs, who agreed to provide a response soon. While OpenAI has yet to receive any updated

6 responses, it does not believe that this dispute requires Court intervention at this time.

7 **Microsoft and GitHub's Position**: On March 5, 2024, Microsoft and GitHub sent a meet and

8 confer letter to Plaintiffs regarding Plaintiffs' deficient responses to GitHub's First Set of Interrogatories.

9 The Parties are still meeting and conferring regarding these responses.   There are no disputes ripe for

10 Court intervention.

11 ## II.   INITIAL DISCLOSURES

12 Pursuant to Federal Rule of Civil Procedure 26(a)(1)(A), Plaintiffs served their Initial Disclosures

13 on June 30, 2023. Defendants[4] served their initial disclosures on July 5, 2023. The Parties will supplement

14 Initial Disclosures, when necessary, as required by Rule 26(e). Plaintiffs have recently sought the

15 disclosure of any indemnity agreements between OpenAI and Microsoft that were not previously

16 disclosed.

17 ## III.   DISCOVERY REFERRAL

18 **Plaintiffs' Position:** On November 21, 2023, this Court issued its Order Referring All Discovery

19 Disputes to Magistrate Judge, which directs the parties to bring all future discovery disputes to Magistrate

20 Judge Ryu according to her procedures. ECF No. 175. Pursuant to Judge Ryu's standing order,[5] the parties

21 must have last met and conferred within five days of bringing any discovery dispute to the Court by joint

22 letter brief. A joint letter brief, ECF No. 213, regarding certain discovery requests related to information

23 about Microsoft's acquisition of GitHub is currently before Judge Ryu, who has postponed ruling on them

24 until after the pleadings have been resolved. ECF No. 226. Because the Court has delayed key discovery

25 until after the pleadings have been resolved and because the Defendants have been uncooperative and

26

27

---

[4] GitHub, Microsoft, and OpenAI are collectively referred to as "Defendants."

28 [5] https://www.cand.uscourts.gov/dmr-revised-standing-order-072022/

1 reticent in producing relevant responsive documents, Plaintiffs anticipate the need will arise to extend

2 discovery in this matter.

3   **OpenAI's Position:** OpenAI disagrees that there is any need to extend discovery at this time, as

4 training data used for Codex can be made available for inspection well before fact discovery closes.

5 Moreover, Plaintiffs have not demonstrated that any additional time would be necessary, particularly when

6 it has not raised any specific disputes with OpenAI's document production or its written discovery

7 responses.

8   **Microsoft and GitHub's Position:** Judge Ryu denied the letter briefs regarding the Acquisition

9 and ordered the Parties to file a new joint discovery letter brief on the issue once the pleadings are settled.

10 ECF No. 226.

11 **IV.**   **CASE MANAGEMENT ORDERS**

12   On August 7, 2023, the Court entered an Amended Stipulated Protective Order For Litigation

13 Involving Patents, Highly Sensitive Confidential Information AND/OR Trade Secrets. ECF No. 145

14 ("Protective Order"). On October 30, 2023, the Court entered a Stipulated Order Concerning Expert

15 Discovery. ECF No. 167 ("Expert Discovery Order"). On November 1, 2023, the Court entered an Order

16 Re Discovery of Electronically Stored Information. ECF No. 168 ("ESI Protocol"). On November 21,

17 2023, the Court ordered all discovery disputes be referred to Magistrate Judge Donna M. Ryu. ECF No.

18 175. On December 12, 2023, the Court ordered the case management conference scheduled for December

19 20, 2023, to be continued to March 22, 2024. ECF No. 186.

20 **V.**   **ISSUES REQUIRING THE COURT'S ATTENTION**

21   **A.**   **Motion for Leave to File a Motion for Reconsideration**

22   Plaintiffs filed a motion for leave to file a motion for reconsideration of the Court's decision on

23 Defendants' Motions to Dismiss the First Amended Complaint on the grounds that the Court had

24 dismissed injunctive relief and damages claims that were not the subject of the prior motion to dismiss and

25 with respect to the "identicality" requirement of DMCA § 1202(b)(1) and (b)(3). ECF No. 218. The Court

26 ordered Defendants to file an opposition to this motion by March 15, 2024. ECF No. 224.

27 The Plaintiffs anticipate that the Court or magistrate Judge Ryu will be asked to determine the

28 applicability of the model patent ESI protocol to this case. This matter is not typical patent litigation and

1   involves far more nuanced issues regarding Defendants' creation of Copilot and Defendants' knowledge

2   of and intent to breach open-source licenses and GitHub's Policies, and to violate the DMCA in doing so.

3   Plaintiffs anticipate that Defendants will oppose any expansion and will seek to impose substantial limits

4   on ESI discovery.

5   **B.   Bifurcation of Classwide Damages Discovery**

6   **Microsoft and GitHub's Position:** Microsoft and GitHub request the Court issue an order setting

7   forth a procedure for addressing bifurcation of discovery regarding classwide damages. Specifically, the

8   Court should permit Plaintiffs to seek such discovery only if Plaintiffs can make a *prima facie* showing for

9   a damages class. *See Mantolete v. Bolger*, 767 F.2d 1416, 1424 (9th Cir. 1985), *as amended* (Aug. 27,

10   1985) (holding that "the plaintiff bears the burden of advancing a *prima facie* showing that the class action

11   requirements of Fed.R.Civ.P. 23 are satisfied or that discovery is likely to produce substantiation of the

12   class allegations" and that "a trial court's refusal to allow class discovery is not an abuse of discretion"

13   absent said "prima facie showing"); *Kaminske v. JP Morgan Chase Bank N.A.*, No. SACV 09-00918 JVS,

14   2010 WL 5782995, at *2 (C.D. Cal. May 21, 2010). While a prima facie showing is not required in all

15   cases, for the following reasons, the Court should exercise its discretion here and require Plaintiffs make

16   such a showing before permitting classwide damages discovery.

17   Bifurcation is necessitated by Plaintiffs' wide-ranging discovery campaign into Microsoft's 2018

18   acquisition of GitHub (the "Acquisition"), which Plaintiffs hope to use as a proxy for class damages. See

19   ECF No. 213 (discovery letter brief) at 4 ("**The acquisition constitutes the most direct valuation by**

20   **Microsoft of the Licensed Materials that belong to Plaintiffs and the Class**." (emphasis in original)).

21   Plaintiffs have served more than ten RFPs and multiple interrogatories on GitHub and Microsoft seeking

22   every document and communication regarding the Acquisition. They have served six third-party

23   subpoenas regarding the Acquisition. There have been seven discovery letter briefs regarding the

24   Acquisition, five of which were exclusively regarding the Acquisition, and two of which dragged third

25   parties (who have nothing to do with Copilot, Codex, or any claim in this case) into discovery disputes.

26   *See* ECF Nos. 181-182, 184, 202-203, 208-209, and 213. Indeed, Plaintiffs appear to be more concerned

27   with obtaining discovery about the Acquisition than any other subject.

28

1   Meanwhile, the prospect of any classwide calculation of damages being proper is remote. The

2 current claims in the case upon which Plaintiffs have stated any claim for damages are claims for breach

3 of contract, where the measure of relief is benefit of the bargain. This is a uniquely individualized

4 determination where the claimed harm is a lack of attribution, or personal reputational credit for authoring

5 source code. Plaintiffs seek wide-ranging discovery into the GitHub Acquisition as a proxy for classwide

6 source code valuation, which is not even an available measure of damages on their remaining claims.

7 Accordingly, Defendants Microsoft and GitHub request that the Court provide a process whereby they are

8 afforded an opportunity to address bifurcation of classwide damages discovery prior to Plaintiffs pursuing

9 any further either party or third-party discovery on this topic.

10   **Plaintiffs' Position:** Microsoft hopes to keep its intent to breach its duties under the open-source

11 licenses and GitHub's Policies shielded from discovery any way possible, including by resisting discovery

12 into its acquisition of GitHub, and how and why it valued Plaintiffs' code repositories. Contrary to

13 Microsoft's repeated attempts to characterize access to information regarding the acquisition as only

14 pertaining to class-wide damages, such information goes directly to Microsoft's understanding and breach

15 of the open-source license terms under which each Plaintiff's code was posted to GitHub, as well as

16 Microsoft's interpretations of GitHub's Policies and promises not to sell each Plaintiff's code. Microsoft's

17 intent in acquiring GitHub to create Copilot, and how and why it valued Plaintiffs' code, is relevant to

18 every individual claim in this matter, regardless of class certification issues.

19   In a further effort to foreclose relevant discovery, Microsoft suggests bifurcating merits and class

20 discovery. In so doing, Microsoft essentially concedes the discovery is proper, but asks to foreclose it until

21 after discovery. Microsoft had an opportunity to raise this when the Court first considered the parties'

22 proposals regarding a pretrial litigation schedule. Had Microsoft done so at the time, Plaintiffs would have

23 had an opportunity to address this issue. Briefly, bifurcation of merits and class discovery is disfavored,

24 particularly in light of Plaintiffs' evidentiary burden at class certification and the fact that merits and class

25 discovery are intertwined. *See Comcast Corp. v. Behrend*, 569 U.S. 27 (2013). The following factors are

26 generally relevant in deciding whether bifurcation is appropriate: (1) the overlap between individual and

27 class discovery; (2) whether bifurcation promotes Rule 23(c)(1)(A)'s requirement that certification be

28 decided "[a]t an early practicable time"; (3) judicial economy; and (4) any prejudice reasonably likely to

1 flow from the grant or denial of a stay of class discovery. *Blair v. Assurance IQ LLC*, No. C23-0016-KKE,

2 2023 WL 6622415, at *6 (W.D. Wash. Oct. 11, 2023). These factors, in addition to the overlap of

3 individual and class issues, weigh heavily against bifurcation.

4   Defendants now want to upset and revise the litigation schedule through a case management

5 conference statement. Defendants already agreed to a litigation schedule that did not involve bifurcation of

6 class issues and already agreed to the set dates for class certification briefing and the Court entered its

7 order on that basis. See ECF No. 130. Such a last-minute request for bifurcation is improper. See e.g.,

8 *True Health Chiropractic Inc v. McKesson Corp.*, No. 13-CV-02219-JST, 2015 WL 273188, at *3 (N.D.

9 Cal. Jan. 20, 2015) ("An additional consideration weighs against bifurcating individual and class

10 discovery in this case—namely, the relatively late date of the request, especially given that the parties

11 initially agreed not to bifurcate discovery.") There is no reason to wreak havoc on the litigation schedule

12 in order to delay or foreclose discovery. *See* Fed. R. Civ. P. 1 ("the Federal Rules of Civil Procedure

13 should be construed, administered, and employed by the court and the parties to secure the just, speedy,

14 and inexpensive determination of every action and proceeding.")

15   **<u>Amendment to Stipulated Order Regarding Telemetry Data Preservation</u>**

16   **GitHub's Position:** On October 10, 2023, the Court granted Microsoft, GitHub, and Plaintiffs'

17 stipulation and proposed order requiring GitHub to maintain until February 21, 2025, in an archived state,

18 telemetry data it would otherwise delete in the ordinary course of business. ECF No. 162. GitHub now

19 requests the Court modify that order to allow GitHub to discontinue preserving newly generated telemetry

20 data. The engineering required to preserve the massive volume of telemetry data has proven to be far more

21 difficult than originally anticipated. The volume of telemetry data will soon surpass GitHub's current

22 technological ability to process, archive, and store the data without engineering new retention systems at

23 great expense. Further, the volume of telemetry data GitHub has preserved to date is more than can

24 possibly be produced or fully queried without sampling. Accordingly, as GitHub has preserved a

25 repository of telemetry data larger than Plaintiffs can possibly query, there is no benefit to Plaintiffs in

26 requiring GitHub to engineer new data stores required to retain additional telemetry data. *See Rodriguez*

27 *v. Google*, No. 20-CV04688-RS (AGT), 2021 WL 8085492, at *1 (N.D. Cal. Dec. 1, 2021) (defendant

28

1   need not maintain data beyond normal retention period when data volume is too large to individually

2   review and adequate amount of data will remain available for sampling).

3   The volume of telemetry data GitHub has collected and preserved for this litigation is enormous

4   and growing rapidly. As of February 2024, GitHub has preserved more than 1 trillion rows of telemetry

5   data, comprising almost 3.5 petabytes. To provide a sense of scale, one petabyte is equivalent to

6   approximately 500 billion pages of standard printed text or 2,000 years of continuous HD video content.

7   In order to preserve this massive amount of data, GitHub engineered an archival data store, that, in

8   October 2023 when it entered into the preservation stipulation, it believed would be able to archive and

9   preserve telemetry data generated through February 21, 2025. That belief has proven to be incorrect. The

10 volume of telemetry data is now starting to overwhelm the data store in terms of both rate and daily

11 capacity, and is expected to hit overall capacity soon. To be able to continue storing the flow of telemetry

12 data, GitHub will need to undertake a significant engineering effort to develop a new space capable of

13 storing the additional data.

14   As things stand, the current data store will not be usable given its size. Running queries on a

15 sample set of the current data is possible, but the engineering effort required to query the full 3.5 petabytes

16 is likely impossible due to technical limitations of search programs, such as timeout periods, maximum

17 size of returnable search results, and limitations on the number of concurrent searches.

18   Accordingly, during a February 13, 2024 in-person meet and confer, GitHub disclosed to Plaintiffs

19 the technical limitations it now faces regarding the ability to continue preserving the volume of new

20 telemetry data and proposed amending the current stipulated order. On February 29, 2024, GitHub wrote

21 to Plaintiffs and reiterated its request that Plaintiffs agree that no further telemetry data need be subject to

22 the original retention stipulation, and that all future telemetry data be subject to ordinary retention policies

23 as indicated by ordinary course business practices. GitHub further stated it would provide a draft

24 amended stipulation and proposed order to address the issue, but that if an agreement with Plaintiffs could

25 not be reached, GitHub would seek relief from the Court in this CMS.

26   On March 4, 2024, GitHub provided Plaintiffs with a draft "Second Stipulation and Amended

27 [Proposed] Order re Preservation of Telemetry Data." The draft proposed that (1) GitHub would maintain

28 in an archived state all telemetry data it has collected through March 1, 2024; (2) all telemetry data

1 generated after March 1, 2024 would be retained in accordance with ordinary business practices; and (3)

2 the 3.5 petabytes of archived data would be preserved through the close of expert discovery. GitHub

3 stated its willingness to confer with Plaintiffs at their convenience, and reiterated that if unable to resolve

4 the issue by agreement, GitHub would raise the issue in this CMS, and if necessary, file a motion.

5   On March 13, 2024, GitHub emailed Plaintiffs again seeking approval of GitHub's proposal.

6 Plaintiffs responded that they see no reason to disturb the existing stipulation as the 3.5 petabytes of data

7 currently preserved is consistent with GitHub's projection during September 2023 meet and confer.

8 Plaintiffs claim to be confused as to what circumstances have changed to warrant departing from the

9 existing stipulation. Plaintiffs' position is that it will consider amending the current preservation

10 stipulation only after the telemetry data store reaches 9 petabytes.

11   What has changed is that the 3.5 petabyte of data is overwhelming the data store GitHub was able

12 to engineer for this case. GitHub will soon not be able to capture the full flow of daily telemetry data and

13 will soon reach the system's overall capacity. Plaintiffs' position that GitHub need preserve 9 petabytes

14 of telemetry data is arbitrary and ludicrous. It is already impossible to fully search the 3.5 petabytes of

15 data that currently exists. As it stands, any queries Plaintiffs request GitHub to perform will have to be

16 performed on a sample of the data. Accordingly, requiring GitHub to engineer new data flows and

17 retention systems to store 9 petabytes, when the 3.5 petabytes of current data is too large to fully search, is

18 disproportionate to Plaintiffs' needs and not based in the reality of computing capabilities. The Court

19 should amend the stipulated order and relieve GitHub of the unreasonable burden of retaining additional

20 telemetry data beyond the massive store it has already preserved.

21   **Plaintiffs' Position:** Nothing has changed since the Court's order of October 10, 2023 (ECF No.

22 162) granting the Parties' stipulation that GitHub and Microsoft would preserve the relevant Telemetry

23 Data in this matter through February 21, 2025. Defendant GitHub is a wholly owned subsidiary of

24 Defendant Microsoft, which hosts CoPilot's operations and data on its own Azure cloud-computing

25 platform. The 3.5 PB of Telemetry Data is consistent with GitHub and Microsoft's own projections of the

26 amount of data it agreed to store before stipulating to store it. On September 27, 2023, Microsoft's and

27 GitHub's counsel stated that by "January 1, 2024 there are projected to be over 2.8 petabytes of active

28 telemetry data. By October 1, 2024 (just after the current fact discovery cut-off date), there are projected

1   to be over 6 petabytes of active telemetry data. By March 1, 2025 (just after the expert discovery cut-off

2   date), there are projected to be over 9 petabytes of active telemetry data." Thus, the amount of data it is

3   currently storing is consistent with its own expectations before agreeing to store it. To the extent the

4   GitHub and Microsoft are running up against their own self-imposed limitations in storage and access to

5   Copilot's Telemetry Data and will be required to address "new data flows and retention systems,"

6   Plaintiffs contend that is not sufficient reason to depart from the Court's previous order.

7   **VI.   ISSUES BEARING ON THE PROGRESS OF THE CASE**

8   **A. OpenAI Global, LLC and Other Unnamed OpenAI-related Entities**

9   **Plaintiffs' Position**: Plaintiffs added certain OpenAI entities to the Second Amended Complaint.

10   ECF No. 200. The Parties are in the middle of briefing Defendants' motions to dismiss the SAC. ECF

11   Nos. 214-2 and 219.

12   **OpenAI's Position**: OpenAI has repeatedly explained to Plaintiffs through email correspondence

13   and during their meet and confers that certain OpenAI entities either do not have any involvement in the

14   claims and defenses at issue in this case or, because of a name change, do not even exist anymore.

15   OpenAI has expressed its willingness to further meet and confer with Plaintiffs regarding which OpenAI

16   entities are actually relevant to the case and to narrow the named OpenAI entities. Plaintiffs have

17   disregarded OpenAI's offer and instead, have opted to include in the Second Amended Complaint all

18   OpenAI entities they are aware of, regardless of whether the particular OpenAI entity even exists, is in any

19   way involved in Codex, or is in possession of or has control over relevant documents.

20   **B. Settlement Conference**

21   The Parties participated in a settlement conference on January 19, 2024, with Magistrate Judge

22   Sallie Kim. (*See* ECF No. 137.)

23

24

25

26

27

28

1   Dated: March 15, 2024   By: */s/ Joseph R. Saveri*
                                                          Joseph R. Saveri

2

3   Joseph R. Saveri (State Bar No. 130064)
                                               Cadio Zirpoli (State Bar No. 179108)
4   Christopher K.L. Young (State Bar No. 318371)
                                               Louis A. Kessler (State Bar No. 243703)
5   Elissa A. Buchanan (State Bar No. 249996)
                                               William W. Castillo Guardado (State Bar No. 294159)
6   Holden J. Benon (State Bar No. 325847)
                                               JOSEPH SAVERI LAW FIRM, LLP
7   601 California Street, Suite 1000
                                               San Francisco, California 94108
8   Telephone: (415) 500-6800
                                               Facsimile: (415) 395-9940
9   Email: jsaveri@saverilawfirm.com
                                                               czirpoli@saverilawfirm.com
10   cyoung@saverilawfirm.com
                                                               lkessler@saverilawfirm.com
11   eabuchanan@saverilawfirm.com
                                                               tmanfredi@saverilawfirm.com
12   wcastillo@saverilawfirm.com
                                                               hbenon@saverilawfirm.com
13
                                               Matthew Butterick (State Bar No. 250953)
14   1920 Hillhurst Avenue, #406
                                               Los Angeles, CA 90027
15   Telephone: (323) 968-2632
                                               Facsimile: (415) 395-9940
16   Email: mb@buttoricklaw.com

17

18   *Counsel for Plaintiffs and the Proposed Class*

19

20

21

22

23

24

25

26

27

28

1   Dated: March 15, 2024

2   By: /s/ Annette L. Hurst
                                                               Annette L. Hurst

3   Annette L. Hurst (State Bar No. 148738)
                                                     ORRICK, HERRINGTON & SUTCLIFFE LLP
4   The Orrick Building
                                                     405 Howard Street
5   San Francisco, CA 94105
                                                     Telephone: 415-773-5700
6   Facsimile: 415-773-5759
                                                     Email: ahurst@orrick.com
7

8   *Counsel for GitHub, Inc. and Microsoft Corporation*

9

10 Dated: March 15, 2024
                                                     By: /s/ Rose Lee
11   Rose Lee

12   Joseph C. Gratz (State Bar No. 240676)
                                                     Rose S. Lee (State Bar No. 294658)
13   **MORRISON & FOERSTER LLP**
                                                     425 Market Street
14   San Francisco, CA 94105
                                                     Telephone: 415-268-7000
15   Facsimile: 415-268-7522
                                                     Email: jgratz@mofo.com
16

17   *Attorneys for Defendants OPENAI, INC., OPENAI,
                                                     L.P., OPENAI OPCO, L.L.C., OPENAI GP, L.L.C.,*
18   *OPENAI STARTUP FUND GP I, L.L.C., OPENAI
                                                     STARTUP FUND I, L.P., OPENAI STARTUP FUND*
19   *MANAGEMENT, LLC, OPENAI, L.L.C., OPENAI
                                                     GLOBAL, LLC, OAI CORPORATION, OPENAI*
20   *HOLDINGS, LLC, OPENAI HOLDCO, LLC, OPENAI
                                                     INVESTMENT LLC, OPENAI STARTUP FUND SPV I,*
21   *L.P., and OPENAI STARTUP FUND SPV GP I, L.L.C.,*

22

23

24

25

26

27

28

1

L.R. 5-1 SIGNATURE ATTESTATION

2   As the ECF user whose user ID and password are utilized in the filing of this document, I attest

3 under penalty of perjury that concurrence in the filing of the document has been obtained from each of the

4 other signatories.

5

6   Dated: March 15, 2024

    /s/ Joseph R. Saveri
7   Joseph R. Saveri

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28